REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
JUNE 2013	CONFERENCE PAPER (Post Print)	JUN 2010 – FEB 2011

4. TITLE AND SUBTITLE

ON THE EXTRACTION OF SPREAD-SPECTRUM HIDDEN DATA IN DIGITAL MEDIA

5a. CONTRACT NUMBER
IN HOUSE

5b. GRANT NUMBER
FA8750-11-1-0016

5c. PROGRAM ELEMENT NUMBER
62702F

6. AUTHOR(S)

M. Kulhandjian, D. Pados, S. Batalama, D. Pados, M. Medley and J. Matyjas

5d. PROJECT NUMBER
ANCL

5e. TASK NUMBER
62

5f. WORK UNIT NUMBER
07

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
State University of New York at Buffalo
Department of Electrical Engineering
332 Bonner Hall
Buffalo, NY 14260-2050

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory/Information Directorate
Rome Research Site/RITE
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TP-2013-027

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA Case Number: 88ABW-2011-3182
DATE CLEARED: 06 JUNE 2011

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This paper considers the problem of blindly extracting data embedded over a wide band in a spectrum (transform) domain of a digital medium (image, audio, video). We first develop a multi-signature iterative generalized least-squares (M-IGLS) core procedure to seek unknown data hidden in hosts via multi-signature direct-sequence spread-spectrum embedding. Neither the original host nor the embedding signatures are assumed available. Then, cross-correlation enhanced M-IGLS (CC-M-IGLS), a procedure described herein in detail that is based on statistical analysis of repeated independent M-IGLS processing of the host, is seen to offer most effective hidden message recovery. Experimental studies on images show that the proposed CC-M-IGLS algorithm can achieve recovery probability of error close to what may be attained with known embedding signatures and host autocorrelation matrix.

15. SUBJECT TERMS
Algorithm design and analysis , Bit error rate , Correlation , Data mining , Reliability , Transforms , Vectors

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			MICHAEL J. MEDLEY
U	U	U	UU	6	19b. TELEPHONE NUMBER (*Include area code*) N/A

On the Extraction of Spread-Spectrum Hidden Data in Digital Media

Ming Li, Michel Kulhandjian, Dimitris A. Pados†, Stella N. Batalama
Department of Electrical Engineering
State University of New York at Buffalo
Buffalo, NY 14260 USA
E-mail: {mingli,mkk6,batalama,pados}@buffalo.edu

Michael J. Medley, John D. Matyjas
Air Force Research Laboratory/RITF
525 Brooks Rd
Rome, NY 13441 USA
E-mail: {michael.medley,john.matyjas}@rl.af.mil

***Abstract*—This paper considers the problem of blindly extracting data embedded over a wide band in a spectrum (transform) domain of a digital medium (image, audio, video). We first develop a multi-signature iterative generalized least-squares (M-IGLS) core procedure to seek unknown data hidden in hosts via multi-signature direct-sequence spread-spectrum embedding. Neither the original host nor the embedding signatures are assumed available. Then, cross-correlation enhanced M-IGLS (CC-M-IGLS), a procedure described herein in detail that is based on statistical analysis of repeated independent M-IGLS processing of the host, is seen to offer most effective hidden message recovery. Experimental studies on images show that the proposed CC-M-IGLS algorithm can achieve recovery probability of error close to what may be attained with known embedding signatures and host autocorrelation matrix.**

Index Terms–Authentication, blind detection, covert communications, data hiding, information hiding, spread-spectrum embedding, steganalysis, steganography, watermarking.

I. Introduction

Digital data embedding in digital media is an information technology field of rapidly growing commercial, as well as national security, interest. Applications of digital data embedding include authentication in its various forms (for example, permanent "iron branding" to show ownership, fragile watermarking to detect future tampering, hidden low-probability-to-detect identification for confidential data validation, etc.) and steganography whose purpose is to establish covert communication between trusting parties. The broad common objective of steganographic applications is a satisfactory tradeoff between hidden data resistance to noise/disturbance (robustness), information delivery rate (payload), and low host distortion for concealment purposes.

The countermeasure technology to data hiding is frequently referred to as steganalysis. Steganalysis can be classified into two categories, *passive* and *active*. The primary task of passive steganalysis is to decide the presence or absence of hidden messages in given media objects [1]. In contrast, active steganalysis refers to the effort of extracting the actual hidden data. While passive steganalysis is being intensively investigated in the past few years, active steganalysis is a relatively new branch of research [2].

In this work, we focus our attention on active spread-spectrum (SS) steganalysis. In particular, we aim to recover blindly data hidden in hosts via (multi-signature) direct-sequence SS embedding [3]-[6]. Neither the original host nor the embedding signatures (spreading sequences) are known (fully blind SS steganalysis). In blind active SS steganalysis the unknown host acts as a source of interference/disturbance to the data to be extracted and, in a way, the problem parallels blind signal separation (BSS) applications as they arise in the fields of array processing, biomedical signal processing, and code-division multiple-access (CDMA) communication systems. Under the assumption that the embedded secret messages are independent identically distributed (i.i.d.) random sequences and independent to the cover host, independent component analysis (ICA) -one particular family of BSS methods- may be utilized to approach the hidden data extraction problem [2],[7]. However, ICA-based BBS algorithms degrade rapidly in the presence of correlated signal interference as is exactly the case in SS image/video/audio embedding. In [8], Gkizeli *et al.* developed an iterative generalized least squares (IGLS) procedure to blindly recover unknown messages hidden in image hosts via SS embedding. The algorithm has low complexity and strong recovery performance. However, the scheme is designed solely for *single-signature* SS embedding where messages are hidden with one signature only and is not generalizable to the *multi-signature* case. Realistically, a steganographer would favor *multi-signature* SS embedding to increase security and payload rate.

In this paper, we develop a new *multi-signature* iterative generalized least squares (M-IGLS) SS steganalysis algorithm for hidden data extraction. For improved recovery performance and in particular for small hidden messages that pose the greatest challenge, we propose an algorithmic upgrade referred to as cross-correlation enhanced M-IGLS (CC-M-IGLS). CC-M-IGLS relies on statistical analysis of independent M-IGLS executions on the host and experimental studies demonstrate hidden data recovery with probability of error close to what may be attained with known embedding signatures and known original host autocorrelation matrix.

II. Multi-signature SS Embedding and Steganalysis Problem Formulation

Consider a host image $\mathbf{H} \in \mathcal{M}^{N_1 \times N_2}$ where $\mathcal{M}$ is the finite image alphabet and $N_1 \times N_2$ is the image size in pixels. Without loss of generality, the image $\mathbf{H}$ is partitioned into M local non-overlapping blocks of size $\frac{N_1 N_2}{M}$. Each block, $\mathbf{H}_1, \mathbf{H}_2, ..., \mathbf{H}_M$, is to carry K hidden information bits

†Corresponding author. Approved for Public Release; distribution unlimited: 88ABW-2011-3182 dated 06 June 2011.

coming -potentially- from K distinct messages. Embedding is performed in a 2-D transform domain $\mathcal{T}$ (such as the discrete cosine transform, a wavelet transform, etc.). After transform calculation and vectorization (for example by conventional zig-zag scanning), we obtain $\mathcal{T}(\mathbf{H}_m) \in \mathbb{R}^{\frac{N_1 N_2}{M}}, m = 1, 2, \ldots, M$. From the transform domain vectors $\mathcal{T}(\mathbf{H}_m)$ we choose a fixed subset of $L \leq \frac{N_1 N_2}{M}$ coefficients (bins) to form the final host vectors $\mathbf{x}(m) \in \mathbb{R}^L$, $m = 1, 2, \ldots, M$. It is common and appropriate to avoid the dc coefficient (if applicable) due to high perceptual sensitivity in changes of the dc value.

A. Multi-signature SS Embedding

The K distinct message bit sequences $\{b_k(m)\}_{m=1}^M$, $k = 1, 2, \ldots, K$, $b_k(m) \in \{\pm 1\}$, are hidden in the transform-domain host vectors $\{\mathbf{x}(m)\}_{m=1}^M$ via additive SS embedding by means of K spreading sequences (signatures) $\mathbf{s}_k \in \mathbb{R}^L, \|\mathbf{s}_k\| = 1$, $k = 1, 2, \ldots, K$,

$$\mathbf{y}(m) = \sum_{k=1}^{K} A_k b_k(m) \mathbf{s}_k + \mathbf{x}(m) + \mathbf{n}(m),\ m = 1, 2, \ldots, M, \quad (1)$$

with corresponding amplitudes $A_k > 0$, $k = 1, \ldots, K$; for the sake of generality, $\mathbf{n}(m) \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \mathbf{I}_L)$ represents potential external white Gaussian noise[1] with variance σ_n^2. It is assumed that $b_k(m)$ behave as equi-probable binary random variables that are independent in time, $m = 1, ..., M$, and across messages, $k = 1, ..., K$. The contribution of each individual embedded message bit b_k to the composite signal is $A_k b_k \mathbf{s}_k$ and the mean-squared distortion to the original host data $\mathbf{x}$ due to the embedded k message alone is

$$\mathcal{D}_k = \mathbb{E}\{\|A_k \mathbf{s}_k b_k\|^2\} = A_k^2,\ \ k = 1, 2, ..., K. \quad (2)$$

The intended recipient of the kth message can perform hidden bit detection by looking at the sign of the output of the minimum-mean-square-error (MMSE) filter $\mathbf{w}_{MMSE,k}$:

$$\hat{b}_k(m) = \mathrm{sgn}\{\mathbf{w}_{MMSE,k}^T \mathbf{y}(m)\} = \mathrm{sgn}\{\mathbf{s}_k^T \mathbf{R}_{\mathrm{y}}^{-1} \mathbf{y}(m)\} \quad (3)$$

where $\mathbf{R}_{\mathrm{y}}$ is the autocorrelation matrix of the stego vectors $\{\mathbf{y}(m)\}_{m=1}^M$

$$\mathbf{R}_{\mathrm{y}} \triangleq \mathbb{E}\{\mathbf{y}\mathbf{y}^T\} = \mathbf{R}_{\mathrm{x}} + \sum_{k=1}^{K} A_k^2 \mathbf{s}_k \mathbf{s}_k^T + \sigma_n^2 \mathbf{I}_L. \quad (4)$$

The autocorrelation matrix $\mathbf{R}_{\mathrm{y}}$ can be estimated by sample averaging over the finite set of M stego data, $\widehat{\mathbf{R}}_{\mathrm{y}} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{y}(m) \mathbf{y}(m)^T$. Using $\widehat{\mathbf{R}}_{\mathrm{y}}$ in (3), we obtain what is known as the sample-matrix-inversion MMSE (SMI-MMSE) detector implementation.

B. Formulation of Active Steganalysis Problem

We assume that the active data extraction analyst has the ability to obtain transform domain stego data in the form of $\mathbf{y}(m)$ in (1) after performing appropriate image partition, transform, and coefficient selection[2] on the image classified as stego by passive steganalysis. We denote the combined "disturbance" to the hidden data (host plus noise) by $\mathbf{z}(m) \triangleq \mathbf{x}(m) + \mathbf{n}(m)$. Then, SS embedding by (1) can be rewritten as

$$\mathbf{y}(m) = \sum_{k=1}^{K} A_k b_k(m) \mathbf{s}_k + \mathbf{z}(m),\ m = 1, \ldots, M, \quad (5)$$

where $\mathbf{z}(m)$ is modeled as a sequence of zero-mean (without loss of generality) vectors with autocovariance matrix $\mathbf{R}_{\mathrm{z}} = \mathbb{E}\{\mathbf{z}\mathbf{z}^T\} = \mathbf{R}_{\mathrm{x}} + \sigma_n^2 \mathbf{I}$. Let $\mathbf{v}_k \triangleq A_k \mathbf{s}_k \in \mathbb{R}^L$, $k = 1, \ldots, K$, be amplitude-including embedding signatures. Then, we can further rewrite SS embedding as

$$\begin{aligned} \mathbf{y}(m) &= \sum_{k=1}^{K} b_k(m) \mathbf{v}_k + \mathbf{z}(m) && (6) \\ &= \mathbf{V}\mathbf{b}(m) + \mathbf{z}(m),\ m = 1, \ldots, M, && (7) \end{aligned}$$

where $\mathbf{V} \triangleq [\mathbf{v}_1, \ldots, \mathbf{v}_K] \in \mathbb{R}^{L \times K}$ is the amplitude-including signature matrix and $\mathbf{b}(m) \in \{\pm 1\}^{K \times 1}$ is the vector of bits embedded in the mth host block. For notational simplicity, we can write the whole stego image data as one matrix

$$\mathbf{Y} = \mathbf{V}\mathbf{B} + \mathbf{Z} \quad (8)$$

where $\mathbf{Y} \triangleq [\mathbf{y}(1)\ \mathbf{y}(2)\ ...\ \mathbf{y}(M)] \in \mathbb{R}^{L \times M}$, $\mathbf{B} \triangleq [\mathbf{b}(1)\ \mathbf{b}(2)\ ...\ \mathbf{b}(M)] \in \{\pm 1\}^{K \times M}$, and $\mathbf{Z} \triangleq [\mathbf{z}(1)\ \mathbf{z}(2)\ ...\ \mathbf{z}(M)] \in \mathbb{R}^{L \times M}$.

Our objective is to blindly extract the unknown hidden data $\mathbf{B}$ from the stego data $\mathbf{Y}$ without prior knowledge of the embedding signatures $\mathbf{s}_k$, and amplitudes A_k, $k = 1, \ldots, K$, in $\mathbf{V} = [A_1 \mathbf{s}_1, \ldots, A_k \mathbf{s}_k]$ or the host itself $\mathbf{x}(1), \ldots, \mathbf{x}(M)$ in $\mathbf{Z} = [\mathbf{x}(1) + \mathbf{n}(1), \ldots, \mathbf{x}(M) + \mathbf{n}(M)]$.

III. Hidden Data Extraction

If $\mathbf{Z}$ were to be modeled as Gaussian distributed, the joint maximum-likelihood (ML) estimator of $\mathbf{V}$ and detector of $\mathbf{B}$ would be

$$\widehat{\mathbf{V}}, \widehat{\mathbf{B}} = \arg \min_{\substack{\mathbf{B} \in \{\pm 1\}^{(K \times M)}, \\ \mathbf{V} \in \mathbb{R}^{L \times K}}} \|\mathbf{R}_{\mathrm{z}}^{-\frac{1}{2}} (\mathbf{Y} - \mathbf{V}\mathbf{B})\|_F^2 \quad (9)$$

where multiplication by $\mathbf{R}_{\mathrm{z}}^{-\frac{1}{2}}$ can be interpreted as prewhitening of the compound observation data. If Gaussianity of $\mathbf{Z}$ is not to be invoked, then (9) is simply referred to as the joint generalized least-squares (GLS) solution.

A. Multi-signature Iterative Generalized Least-Squares Procedure

In any case, regretfully, joint estimation of $\mathbf{V}$ and detection of $\mathbf{B}$ by (9) has complexity exponential in KM. To manage the computational complexity, we attempt to reach a quality approximation of the solution of (9) by alternating generalized least-squares estimates of $\mathbf{V}$ and $\mathbf{B}$, iteratively, as described below.

Pretend $\mathbf{B}$ is known; the generalized least-squares estimate of $\mathbf{V}$ is

$$\begin{aligned} \widehat{\mathbf{V}}_{\mathrm{GLS}} &= \arg \min_{\mathbf{V} \in \mathbb{R}^{L \times K}} \|\mathbf{R}_{\mathrm{z}}^{-\frac{1}{2}} (\mathbf{Y} - \mathbf{V}\mathbf{B})\|_F^2 \\ &= \mathbf{Y}\mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1}. \quad (10) \end{aligned}$$

[1] Additive white Gaussian noise is frequently viewed as a suitable model for quantization errors, channel transmission disturbances, and/or image processing attacks.

[2] Host image partition may be estimated by examining the difference between neighboring pixels [9]. For each investigated transform, all coefficients (except the dc value) may be considered.

Pretend, in turn, that $\mathbf{V}$ is known; then, the least-squares estimate of $\mathbf{B}$ *over the real field* is

$$\begin{aligned}\widehat{\mathbf{B}}_{\rm GLS}^{\rm real} &= \arg\min_{\mathbf{B}\in\mathbb{R}^{K\times M}} \|\mathbf{R}_{\rm z}^{-\frac{1}{2}}(\mathbf{Y}-\mathbf{V}\mathbf{B})\|_F^2 \\ &= (\mathbf{V}^T\mathbf{R}_{\rm z}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{R}_{\rm z}^{-1}\mathbf{Y}. \qquad (11)\end{aligned}$$

Observing that

$$(\mathbf{V}^T\mathbf{R}_{\rm z}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{R}_{\rm z}^{-1} = (\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}, \qquad (12)$$

we rewrite

$$\widehat{\mathbf{B}}_{\rm GLS}^{\rm real} = (\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}\mathbf{Y} \qquad (13)$$

and suggest the approximate binary message solution

$$\begin{aligned}\widehat{\mathbf{B}}_{\rm GLS}^{\rm binary} &= \arg\min_{\mathbf{B}\in\{\pm1\}^{K\times M}} \|\mathbf{R}_{\rm z}^{-\frac{1}{2}}(\mathbf{Y}-\mathbf{V}\mathbf{B})\|_F^2 \\ &\simeq \operatorname{sgn}\{(\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}\mathbf{V})^{-1}\mathbf{V}^T\mathbf{R}_{\rm y}^{-1}\mathbf{Y}\}. \qquad (14)\end{aligned}$$

The proofs of (10), (11), and (12) are omitted due to lack of space.

The *multi-signature iterative generalized least-squares* (M-IGLS) procedure suggested by the two equations (10) and (14) is now straightforward. Initialize $\widehat{\mathbf{B}}$ arbitrarily and alternate iteratively between (10) and (14) to obtain at each step conditionally generalized least squares estimates of one matrix parameter given the other. Stop when convergence is observed. Notice that (14) requires knowledge of the autocorrelation matrix of the stego data $\mathbf{R}_{\rm y}$ which can be estimated by sample averaging over the received data observations, $\widehat{\mathbf{R}}_{\rm y} = \frac{1}{M}\sum_{m=1}^{M}\mathbf{y}(m)\mathbf{y}(m)^T$. The M-IGLS SS steganalysis algorithm is summarized in Table I. Superscripts denote iteration index. For the sake of mathematical accuracy, we emphasize that there is always a global message sign/phase ambiguity present when one considers joint data extraction and signature identification (i.e. for each whole extracted message vector, $i = 1,\ldots,K$, is it $\widehat{\mathbf{b}}_i$ or $-\widehat{\mathbf{b}}_i$?) The sign ambiguity problem can be overcome with a few known or guessed data symbols for sign correction.

B. Cross-Correlation Enhanced M-IGLS

We understand that, with arbitrary initialization, convergence of the M-IGLS procedure described in Table I to the optimal GLS solution of (9) is not guaranteed in general. Extensive experimentation with the algorithm in Table I indicates that, for sufficiently long messages hidden by each signature ($M = 4$Kbits or more, for example), satisfactory quality message decisions $\widehat{\mathbf{B}}$ can be obtained. However, when the message size is small, M-IGLS may very well converge/return wrong solutions. The quality (generalized-least-squares fit) of the end convergence point depends heavily on the initialization point and arbitrary initialization -which at first sight is unavoidable for blind steganalysis- offers little assurance that the iterative scheme will lead us to appropriate, "reliable" (close to minimal generalized least-squares fit) solutions. Re-initialization and re-execution of the M-IGLS procedure is always possible but the challenge is how to assess whether solutions returned by the M-IGLS procedure are reliable or not without any side information. The rest of this section is devoted to addressing this challenge.

TABLE I

ITERATIVE GENERALIZED LEAST-SQUARES SS STEGANALYSIS

1) $d := 0$; initialize $\widehat{\mathbf{B}}^{(0)} \in \{\pm1\}^{K\times M}$ arbitrarily.
2) $d := d+1$;
$\widehat{\mathbf{V}}^{(d)} := \mathbf{Y}(\widehat{\mathbf{B}}^{(d-1)})^T\left[(\widehat{\mathbf{B}}^{(d-1)})(\widehat{\mathbf{B}}^{(d-1)})^T\right]^{-1}$;
$\widehat{\mathbf{B}}^{(d)} := \operatorname{sign}\left\{\left((\widehat{\mathbf{V}}^{(d)})^T\widehat{\mathbf{R}}_{\rm y}^{-1}(\widehat{\mathbf{V}}^{(d)})\right)^{-1}(\widehat{\mathbf{V}}^{(d)})^T\widehat{\mathbf{R}}_{\rm y}^{-1}\mathbf{Y}\right\}$.
3) Repeat Step 2 until $\widehat{\mathbf{B}}^{(d)} = \widehat{\mathbf{B}}^{(d-1)}$.

Since $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{V}}$ are jointly detected and estimated, correspondingly, if one is not reliable neither is the other in general. We first examine the reliability of the bit matrix decision $\widehat{\mathbf{B}} = [\widehat{\mathbf{b}}_1,\ldots,\widehat{\mathbf{b}}_K]^T$ returned by the M-IGLS procedure of Table I. The sample cross-correlation between any two bit streams is

$$\eta_{i,j} \triangleq \widehat{\mathbf{b}}_i^T\widehat{\mathbf{b}}_j/M,\ i\neq j,\ i,j=1,\ldots,K. \qquad (15)$$

Formally, the true information bits are independent within user streams and across users. If $\eta_{i,j}$ were to be viewed as approximately normally distributed with zero mean and variance $\frac{1}{M}$, then the probability of $|\eta_{i,j}|$, $i\neq j$, being larger than, say, the threshold value $\frac{3}{\sqrt{M}}$ is very low at about 0.3% (we can calculate $\Pr(|\eta_{i,j}| > \frac{3}{\sqrt{M}}) \approx 0.003$). Motivated by this calculation, we introduce below Criterion 1 that classifies convergence points of the M-IGLS procedure in Table I as "compliant" or not based on the sample statistics of the returned data matrix $\widehat{\mathbf{B}}$.

Criterion 1: If $|\eta_{i,j}| \leq \frac{3}{\sqrt{M}}$ for all $i\neq j \in \{1,2,\ldots,K\}$, then $(\widehat{\mathbf{B}}, \widehat{\mathbf{V}})$ returned by the M-IGLS procedure in Table I are classified as "*Criterion-1-compliant*." ■

Criterion 1 provides the means for coarse identification of unreliable solutions. An unreliable convergence point would then trigger re-initialization and re-execution of the M-IGLS procedure in Table I until a Criterion-1-compliant point is obtained. To enhance the end accuracy of blind hidden data extraction, we propose one additional criterion based on the returned estimated signature matrix $\widehat{\mathbf{V}}$. We will motivate our proposal by examining experimentally the normalized cross-correlation between the estimated signatures $\widehat{\mathbf{v}}_k$ returned by the Criterion-1-equipped M-IGLS procedure and the true signatures $\mathbf{v}_k$, $k = 1,\ldots,K$. We consider as a host example the gray scale 256×256 "Baboon" image perform 8×8 block DCT embedding by (1) over all bins except the dc coefficient with $K = 4$ distinct arbitrary signatures $\mathbf{s}_k \in \mathbb{R}^{63}$ and per-message distortion $\mathcal{D}_k = 31.5$dB, $k = 1,\ldots,4$. For the sake of generality, we also incorporate white Gaussian noise of variance $\sigma_n^2 =$ 3dB. We run the Criterion-1-equipped M-IGLS procedure 400 times. The histogram of the normalized cross-correlation values $\theta_k \triangleq \frac{\widehat{\mathbf{v}}_k^T\mathbf{v}_k}{\|\widehat{\mathbf{v}}_k\|\|\mathbf{v}_k\|}$ of the four hundred returned solutions for message $k = 1$ in Fig. 1 (representative of all other messages) reveals that Criterion 1 is not by itself sufficient to eliminate erroneous solutions. Yet, there

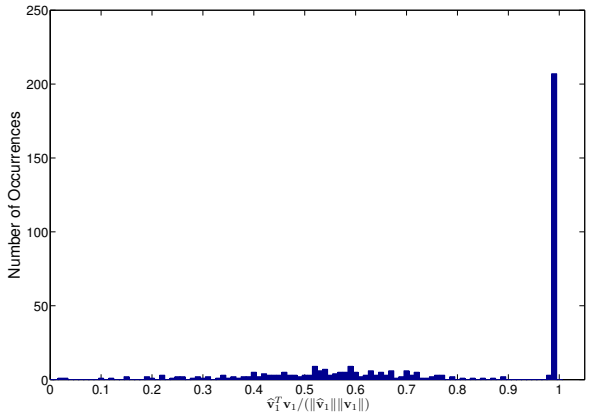

Fig. 1. Histogram of normalized cross-correlation between $\widehat{\mathbf{v}}_1$ and $\mathbf{v}_1$ (256×256 Baboon image, 8×8 DCT, $L=63$, $K=4$, $\mathcal{D}_k=31.5$dB, $k=1,\ldots,4$, $\sigma_n^2=3$dB; $\widehat{\mathbf{v}}_1$ returned by Table I M-IGLS steganalysis procedure).

exists a tight cluster/region formed by 210 or so of the Criterion-1-equipped M-IGLS convergence points around the true embedding signature.

The basic idea now behind our second and final refinement of the M-IGLS blind hidden data extraction procedure is to identify and average these reliable clustered estimates. Of course, identification of the reliable estimates is not a trivial task due to our complete lack of knowledge of $\mathbf{v}_k$ (or $\mathbf{s}_k$), $k=1,\ldots,K$. In this context, assume that we have P estimates of $\mathbf{v}_k$ denoted by $\widehat{\mathbf{v}}_k^{(j)}$, $k=1,\ldots,K$, $j=1,\ldots,P$, obtained by P runs of the Criterion-1-equipped M-IGLS procedure. From the example of Fig. 1, we understand that reliable estimates $\widehat{\mathbf{v}}_k^{(j)}$ of $\mathbf{v}_k$ have high normalized cross-correlation (close to 1) with each other, while they will have low normalized cross-correlation with other unreliable estimates of $\mathbf{v}_k$. In contrast, unreliable estimates will tend to have low normalized cross-correlation with each other. Therefore, the reliability of $\widehat{\mathbf{v}}_k^{(j)}$ may be quantified/assessed by examining the sum-cross-correlation with the other $\widehat{\mathbf{v}}_k^{(t)}$, $t\neq j\in\{1,\ldots,P\}$,

$$\rho_k^{(j)} \triangleq \sum_{t=1,t\neq j}^{P} \frac{|\widehat{\mathbf{v}}_k^{(j)H}\widehat{\mathbf{v}}_k^{(t)}|}{\|\widehat{\mathbf{v}}_k^{(j)}\|\|\widehat{\mathbf{v}}_k^{(t)}\|}\ . \tag{16}$$

A reasonable threshold value for binary reliability classification may be the average value

$$\overline{\rho}_k \triangleq \frac{1}{P}\sum_{j=1}^{P}\rho_k^{(j)},\ k=1,\ldots,K, \tag{17}$$

utilized in the proposed Criterion 2 below.

Criterion 2: Let $\widehat{\mathbf{v}}_k^{(j)}$ be the estimates of $\mathbf{v}_k$ returned by P arbitrary initializations of the Criterion-1-equipped M-IGLS procedure of Table I, $k=1,\ldots,K$, $j=1,\ldots,P$. If $\rho_k^{(j)}\geq\overline{\rho}_k$, then $\widehat{\mathbf{v}}_k^{(j)}$ is considered a *reliable* estimate of the $\mathbf{v}_k$; otherwise we declare it as *unreliable*. ■

Finally, we average our reliable (according to *Criterion 2*) estimates of the effective signatures $\mathbf{v}_k$ to produce one last high-quality initialization of the M-IGLS algorithm of Table I. Let $\mathcal{S}_k$ denote the set of all reliable estimates of $\mathbf{v}_k$ according to *Criterion 2* and let $|\mathcal{S}_k|$ denote the cardinality of $\mathcal{S}_k$. Our averaged estimate of matrix $\mathbf{V}$ is now given by $\overline{\widehat{\mathbf{V}}}$ with

$$\overline{\widehat{\mathbf{V}}} \triangleq \left[\overline{\widehat{\mathbf{v}}}_1,\ldots,\overline{\widehat{\mathbf{v}}}_K\right] \ \text{ where } \ \overline{\widehat{\mathbf{v}}}_k = \frac{1}{|\mathcal{S}_k|}\sum_{j\in\mathcal{S}_k}\widehat{\mathbf{v}}_k^{(j)}\ ,\ k=1\ldots,K, \tag{18}$$

i.e. $\overline{\widehat{\mathbf{v}}}_k$ is the average over all reliable estimates of $\mathbf{v}_k$ according to *Criterion 2*. We execute M-IGLS in Table I a final time initialized at $\widehat{\mathbf{B}}^{(0)}=\text{sgn}\left\{\left(\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\text{y}}^{-1}\overline{\widehat{\mathbf{V}}}\right)^{-1}\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\text{y}}^{-1}\mathbf{Y}\right\}$. We call M-IGLS with both Criteria 1 and 2 incorporated, Cross-Correlation enhanced M-IGLS (CC-M-IGLS) and summarize the complete procedure in Table II.

TABLE II
CROSS-CORRELATION ENHANCED M-IGLS

For $j:=1$ to P
1) Execute M-IGLS of Table I with arbitrary initialization and obtain estimates $\widehat{\mathbf{v}}_k$, $k=1,\ldots,K$.
2) **If** estimates are Criterion-1-compliant, $\widehat{\mathbf{v}}_k^{(j)}:=\widehat{\mathbf{v}}_k$, $k=1,\ldots,K$; **else** go to 1).
End
For $k:=1$ to K
3) Identify reliable estimates for $\mathbf{v}_k$ according to *Criterion 2*.
4) Calculate the average over all reliable estimates $\overline{\widehat{\mathbf{v}}}_k$ by (18).
End
5) Set $\overline{\widehat{\mathbf{V}}}\triangleq[\overline{\widehat{\mathbf{v}}}_1,\ldots,\overline{\widehat{\mathbf{v}}}_K]$.
6) Execute M-IGLS of Table I with initialization $\widehat{\mathbf{B}}^{(0)}=\text{sgn}\left\{\left(\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\text{y}}^{-1}\overline{\widehat{\mathbf{V}}}\right)^{-1}\overline{\widehat{\mathbf{V}}}^T\widehat{\mathbf{R}}_{\text{y}}^{-1}\mathbf{Y}\right\}$.

IV. EXPERIMENTAL STUDIES

A technically firm and keen measure of quality of an active steganalysis solution is the difference in the bit-error-rate (BER) experienced by the intended recipient and the steganalyst. The intended recipient in our studies may be using any of the following three message recovery methods: (*i*) Standard signature matched-filtering (MF) with known signatures $\mathbf{s}_k$, $k=1,...,K$, (*ii*) sample-matrix-inversion MMSE (SMI-MMSE) filtering with known signatures $\mathbf{s}_k$ and estimated host autocorrelation matrix $\widehat{\mathbf{R}}_{\text{y}}$ (see (3)); (*iii*) ideal MMSE filtering with known signatures $\mathbf{s}_k$ and known true host autocorrelation matrix $\mathbf{R}_{\text{x}}$ which serves as the ultimate performance bound reference for all methods. In terms of blind active steganalysis (neither $\mathbf{s}_k$ nor $\mathbf{R}_{\text{x}}$ known), we will examine (*iv*) the developed M-IGLS algorithm in Table I alone and (*v*) CC-M-IGLS of Table II with $P=20$ Criterion-1 runs. Finally, the performance of two typical ICA-based blind signal separation (BSS) algorithms, (*vi*) FastICA [13], and (*vii*) JADE [14], will also be included in the studies for comparison purposes.

We consider as a host example the gray-scale 512×512 "Baboon" image. We perform 8×8 block DCT embedding by (1) over all bins except the dc coefficient with $K=4$ distinct arbitrary signatures $\mathbf{s}_k\in\mathbb{R}^{63}$, $k=1,\ldots,K$. The

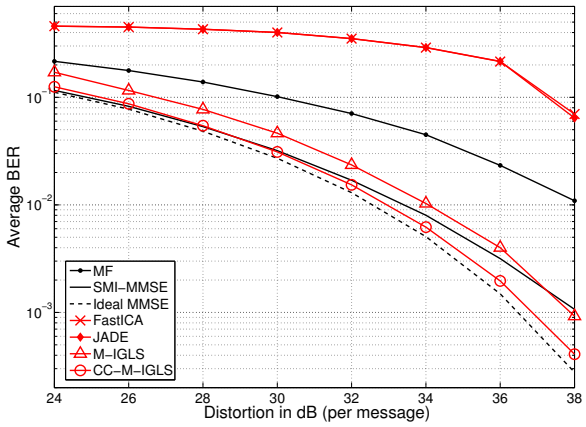

Fig. 2. Average BER versus per-message distortion (512×512 Baboon, $L = 63$, $K = 4$ messages of 4Kbits each, $\sigma_n^2 = 3$dB).

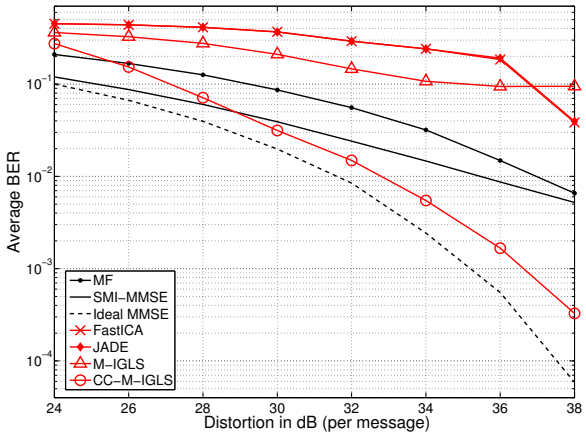

Fig. 3. Average BER versus per-message distortion (256×256 Baboon, $L = 63$, $K = 4$ messages of 1Kbit each, $\sigma_n^2 = 3$dB).

hidden message embedded by each signature is $\frac{512^2}{8^2} = 4,096$ bits long. The per-message mean square distortion due to each embedded message is set to be the same for all messages, i.e. $\mathcal{D}_k = A_k^2 = \frac{\mathcal{D}}{K}$, $k = 1, \ldots, 4$. For the sake of generality, we also incorporate white Gaussian noise of variance $\sigma_n^2 = 3$dB. Fig. 2 shows the average BER (over all $K = 4$ messages) of all methods (*i*) through (*vii*) listed above as a function of the host distortion per message. While the independent/principal-component methods (FastICA and JADE) are failing to carry out effective active SS image steganalysis, to our satisfaction CC-M-IGLS SS steganalysis is rather close in BER performance to the ideal MMSE detector bound for which the embedding signatures and the clean host autocorrelation matrix $\mathbf{R}_x$ are perfectly known. It could be argued that for this host and rather large size of $M = 4,096$ bits per message, CC-M-IGLS offers only a moderate gain in comparison with M-IGLS of Table I by itself.

In Fig. 3, however, we repeat the exact same experimental study on the smaller 256×256 version of the Baboon image with $K = 4$ hidden messages of length only $\frac{256^2}{8^2} = 1,024$ bits per message. CC-M-IGLS now provides dramatic performance improvement over M-IGLS which would justify the extra computational cost and extraction delay. At the same time, comparing with Fig. 2, the gap between CC-M-IGLS and ideal MMSE increases as the hidden message size decreases.

V. Conclusions

In this paper we considered the problem of recovering unknown messages hidden in digital media hosts via multi-signature spread-spectrum embedding. Neither the original host nor the embedding signatures are assumed available. We first developed a low complexity multi-signature iterative generalized least-squares (M-IGLS) core algorithm. Cross-correlation enhanced M-IGLS (CC-M-IGLS), a procedure based on statistical analysis of repeated independent M-IGLS processing of the host, is seen to offer most effective blind hidden message recovery and presents itself as an effective countermeasure to conventional[3] SS data hiding.

References

[1] M. Li, M. Kulhandjian, D. A. Pados, S. N. Batalama, and M. J. Medley, "Passive spread-spectrum steganalysis," in *Proc. IEEE Intern. Conf. Image Proc. (ICIP)*, Brussels, Belgium, Sept. 2011.

[2] R. Chandramouli, "A mathematical framework for active steganalysis," *ACM Multimedia Systems Special Issue on Multimedia Watermarking*, vol. 9, pp. 303-311, Sept. 2003.

[3] H. S. Malvar and D. A. Florencio, "Improved spread spectrum: A new modulation technique for robust watermarking," *IEEE Trans. Signal Proc.*, vol. 51, pp. 898-905, April 2003.

[4] I. J. Cox, J. Kilian, F. T. Leighton, and T. Shannon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Proc.*, vol. 6, pp. 1673-1687, Dec. 1997.

[5] M. Gkizeli, D. A. Pados, and M. J. Medley, "Optimal signature design for spread-spectrum steganography," *IEEE Trans. Image Proc.*, vol. 16, pp. 391-405, Feb. 2007.

[6] M. Gkizeli, D. A. Pados, and M. J. Medley, "SINR, bit error rate, and Shannon capacity optimized spread-spectrum steganography," in *Proc. IEEE Intern. Conf. Image Proc. (ICIP)*, Singapore, Oct. 2004, vol. 2, pp. 1561-1564.

[7] P. Bas and F. Cayre, "Achieving subspace or key security for WOA using natural or circular watermarking," in *Proc. ACM Multimedia and Security Workshop*, Geneva, Switzerland, Sept. 2006.

[8] M. Gkizeli, D. A. Pados, S. N. Batalama, and M. J. Medley, "Blind iterative recovery of spread-spectrum steganographic messages," in *Proc. IEEE Intern. Conf. Image Proc. (ICIP)*, Genova, Italy, Sept. 2005, vol. 2, pp. 11-14.

[9] Y. Wang and P. Moulin, "Steganalysis of block-DCT steganography," in *Proc. IEEE Workshop on Stat. Signal Proc.*, St. Louis, MO, Sept. 2003, pp. 339-342.

[10] S. Talwar, M. Viberg, and A. Paulraj, "Blind seperation of synchronous co-channel digital signals using an antenna array - part I: Algorithms," *IEEE Trans. Signal Proc.*, vol. 44, pp. 1184-1197, May 1996.

[11] M. Li, S. N. Batalama, D. A. Pados, and J. D. Matyjas, "Multiuser CDMA signal extraction," in *Proc. IEEE Military Comm. Conf. (MILCOM)*, Washington D.C., Oct. 2006.

[12] M. Li, S. N. Batalama, and D. A. Pados, "Population size identification for CDMA eavesdropping," in *Proc. IEEE Military Comm. Conf. (MILCOM)*, Orlando, FL., Oct. 2007.

[13] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, pp. 1483-1492, Oct. 1997.

[14] J. F. Cardoso, "High-order contrasts for independent component analysis," *Neural Computation*, vol. 11, pp. 157-192, Jan. 1999.

[3] In [7], Bas and Cayre present an interesting signature-based additive embedding approach different to (1) that is host-vector-by-host-vector dependent and would withstand IGLS-based active steganalysis. The embedding is, however, very sensitive to noise which leads to high recovery error rates by intended recipients and limits the applicability to general covert communication problems.